79-2-11 AUGUST 1979
3430

A TEST OF THE KROLL AND REEVES REGRESSION MODEL FOR PREDICTING NUMBERS OF SPB INFESTATIONS IN ALABAMA AND LOUISIANA

by

J. D. Smith[1/] and G. Ryan[2/]

INTRODUCTION

A model which will predict the number of southern pine beetle SPB spots to be expected in the future is needed to increase the accuracy of SPB evaluations. An effort to build a regression model based on weather variables was attempted and published by authors Kroll and Reeves of Stephen F. Austin University in 1978. In order to test this model for use in other States, the authors of this paper gathered the data required for the Kroll and Reeves regression equation and submitted it to regression analysis for fit. This model used four easily-obtainable weather variables to predict numbers of SPB spots (NSPOTS). These variables are: 1) mean temperature (^{o}F) for February of current year (x_1), 2) total rainfall (in.) for summer of previous year (x_2), 3) total rainfall (in.) for fall of previous year (x_3), and 4) total rainfall (in.) for spring of previous year (x_4). For 2 years the model was effective in east Texas. Numbers of spots found in 1976 and 1977 in east Texas verified the accuracy of the model under epidemic situations.

However, in 1978 the SPB population decreased to endemic or near endemic levels. The prediction capability of the model decreased as the population fell below epidemic levels. The model did not predict accurately the numbers of spots found in Texas in 1978 (38 vs 4100 approximately predicted)(Billings 1978). Data from Louisiana (years 1974 to 1977) and from Alabama (years 1972 to 1977) also could not be used to accurately predict numbers of SPB spots.

1/ Entomologist, Southeastern Area, State & Private Forestry, Forest Insect and Disease Management Group, Pineville, La.

2/ Statistician, Southeastern Area, Forest Insect and Disease Management Staff Unit, Atlanta, Ga.

Four years of data were gathered from Louisiana records for the purpose of analysis. An analysis of variance was run on the data with year significant at the 5 percent level and district significant at the 1 percent level of probability. This suggests that each of these factors should be used in a regression model.

Twelve different models were considered in stepwise regression analysis. In all models, the dependent variable is number of spots (NSPOTS).

Model 2, which includes year, district, and x_2 has the highest R-Square (.3663). Model 1 (year, x_1 and x_2) an R-Square of .3647. There is little reason to prefer one over the other. District can be replaced by x_1 (model 2 vice model 1) because these two variables are highly correlated (r = .486) when compared with other correlation coefficients in this study. So x_1 is a proxy for district, and vice versa. District would be the best variable to use since it is easier to record.

X_3 does not appear in any of the models because it is highly correlated with x_2 (.582). (This is not a high correlation coefficient--this is in comparison to other coefficients found in the analysis of data). X_3 is not needed in the models due to this condition.

X_4 is not correlated with any independent variable or number of spots. Hence, it was not included in the final models. X_1 and x_2 are not correlated with each other, but are correlated with number of spots and therefore are included in the final models.

Models 1 and 2 are the best models. F-ratios for each are significant at the 5 percent level. Although models 1 and 2 are judged to be the best models, the standard error terms are rather large when compared to the mean value for NSPOTS. In summary, none of the models could predict NSPOTS with an acceptable degree of accuracy.

The State of Alabama also gathered data that corresponded to the four variables in Kroll and Reeves paper (x_1, x_2, x_3, x_4). Again, mean squared error was high in all regression runs attempted (120 runs were made integrating $x_1.....x_4$, year, weather service district (WSD), and all years and WSD's combined). Only one model seemed appropriate for use. This was for the year 1974 and had a mean squared error of 350 spots. While 350 spots is not an acceptable level of error, it should be noted that 1974 was an epidemic year for SPB in Alabama. An error term of 350 spots out of a total of approximately 4,000 spots does not seem to be significant. But the error term is greater for endemic years; therefore, the model does not predict trends of activity. This

would seem to indicate that the original model could be adapted to measure fairly well where SPB's are in epidemic numbers. The entomologist would still have to determine whether or not an epidemic was at hand before he/she could use the model. Other years with lower SPB populations were not predicted by the original model or by any adjustments made by adding year, WSD (weather service district), or both.

There is not much hope that any of the models attempted can be used as an effective predictor of NSPOTS. There is a chance that the model could be improved by adding some factors concerning population dynamics of the SPB to it, but more research would need to be done before an attempt could be made.

Suggestions for improving the model are: 1) record several years data from spot growth analysis in each State, combine these variables with the $x_1 \ldots x_4$ variables found by Kroll and Reeves and build a model for the State concerned; 2) record attack/emergence or other population dynamics data and add to the model. Again, this could probably be done best by setting up a model for each State separately.